Salhani, Mohamad; Liinaharja, Markku

# Load migration mechanism in ultra-dense networks

*Published in:*
ICTCE 2018 - Proceedings of the 2018 2nd International Conference on Telecommunications and Communication Engineering

*DOI:*
10.1145/3291842.3291923

Published: 28/11/2018

*Document Version*
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

*Please cite the original version:*
Salhani, M., & Liinaharja, M. (2018). Load migration mechanism in ultra-dense networks. In *ICTCE 2018 - Proceedings of the 2018 2nd International Conference on Telecommunications and Communication Engineering* (pp. 268-274). ACM. https://doi.org/10.1145/3291842.3291923

# Load Migration Mechanism in Ultra-Dense Networks

Mohamad Salhani
COMNET Department,
Aalto University, Finland
mohamad.salhani@aalto.fi

Markku Liinaharja
COMNET Department,
Aalto University, Finland
markku.liinaharja@gmail.com

## ABSTRACT

As one of the main technologies in 5G networks, Ultra-dense networks (UDNs) can be used to improve the network coverage. The dense deployment of small cells in UDN hotspots generates an uneven traffic distribution. In this paper, we propose a novel mechanism in order to transfer the extra users from the small cells to the macrocells based on several load balancing approaches implemented within the small cells, which are formed based on the Radio-over-Fiber (RoF) system. To select the best overlapping zone and then the best candidate user to be handed-over between the access points of the small cells, a common zone approach, a worst zone approach and a mixed approach are proposed. With the objective of transferring the extra users to the macrocells, we suggest a transfer after approach, a transfer before approach and an active approach. The simulation results indicate that the proposed approaches succeed to balance the load among the access points and to migrate the required load from the overloaded small cells to the macrocells in selective way. In some cases, the balance improvement ratio can reach 97.94%. Moreover, the overall balance efficiency is increased by 51.32% compared to the case without transferring users to the macrocells.

## CCS Concepts

**• Network →Network Performance evaluation →Network performance analysis.**

## Keywords

UDN, RoF, load balancing algorithm, common zone approach, worst zone approach, mixed approach, transfer after approach, transfer before approach, active approach.

## 1. INTRODUCTION

The rapid growth of traffic in the coming years will cause macrocell networks to evolve, becoming more tightly packed and eventually ultra-dense. With the bandwidth requirements, the spectrum resources are not sufficient to meet the increased needs such as coverage, speed, latency, etc. The integration of the optical communications and the broadband wireless access technology combines the advantages of optical fiber communications and wireless mobile communications by using UDN networks. Based on the Radio-over-Fiber (RoF) system to form the small cells, the UDN network can improve the coverage and the capacity, and provides higher data-rate transmissions. However, the RoF system suffers from load unbalance [1]. In fact, as the coverage of the RoF system is relatively small, the frequent handovers of the users will influence the system's performance [2]. Hence, a load balancing algorithm (LBA) becomes a necessity to redistribute the load among the access points (APs) of the UDN network and transfer the extra users to the macrocells of the base stations (BSs) in selective way. In studies that have applied the RoF system, the number of handovers was found to be larger than in those discussing traditional cellular networks [3]. Due to the small coverage area of the APs, decreasing handover procedures by optimally selecting the best candidate users can be a promising solution for UDN networks that are based on an RoF system. The first load balancing algorithms within wireless networks were proposed by Balachandran and Aleo [4][5]. These were based on load migration between the APs and can only switch the users of overlapping zones. Nevertheless, the proposed algorithms were very simple and only balanced the load between two cells with an overlapping zone. A channel borrowing scheme has been used to offload the overloaded cells by using unused channel from the neighboring unloaded cells in [6][7]. This method without a strict channel locking strategy may result in co-channel interference. Strategies based on cell breathing and power control have been presented by Hanly *et al* [8]. These can offload the overloaded cells by simultaneously reducing the power of the APs in the overloaded cells and increasing the power of the APs in the underloaded cells. However, these can cause a disconnection of some users located on the cell edges and increase the possibility of co-channel interference, and the AP can remain overloaded even after reducing the coverage area. A new load balancing algorithm in UDN networks based on a stochastic differential game scheme and an RoF system without any policy for transferring the extra users to the macrocells was suggested in [1][2]. Moreover, a QoS constraint optimal load balancing scheme for heterogeneous ultra-dense networks was proposed in [9]. Conversely, these studies did not address the optimization issue of the overlapping zone selection and the user transfer.

To the best of our knowledge, the load balancing by transferring users has not been highlighted enough in the recent studies. Elgendi *et al* [10] have proposed new schemes to find the optimal number of sessions to be transferred from unlicensed long term evolution (U-LTE) networks to licensed long term evolution (L-LTE) or Wi-Fi networks. They have shown that it is possible to transfer the users from programmable base stations to Wi-Fi APs in order to achieve a win-win outcome for both networks. Nonetheless, they have focused on the speed of users and the distance between the user and the BS more than the data offloading. Besides, the proposed schemes have transferred a higher number of users. In this paper, we propose new (user) transfer approaches to offload the small cells of UDN networks by transferring the best candidate users to the macrocells. The transfer approaches are based on several load balancing approaches within the small cells. We first determine the best overlapping zone among many overlapping zones and then the best candidate user to be handed over to another AP or transferred to the BS by selective way in order to reduce the number of the handed over and transferred users, and improve the performance of the whole UDN network. This paper focuses on UDN hotspots where all the APs of the UDN network are considered to be always active. The user density can be 10 times larger than that of APs cells [11]. Hence, the user density can reach six users per each small cell. We assume that the user applications ensure that the throughput of the handed-over user remains constant and the capacity of the BS is always available for the transferred users.

The rest of this paper is organized as follows. The proposed system model is described in Section II. The simulation model and the performance evaluation criteria are illustrated in Sections III and IV. While Section V presents the different load balancing and user transfer algorithms, Section VI introduces the approaches of the load balancing and the user transfer. After that, the results are analyzed in Section VII. Finally, a conclusion and perspectives of this work are presented in Section VIII.

## 2. SYSTEM MODEL

The proposed system consists of multiple macrocells, as shown in Figure 1. Several APs of UDN small cells with overlapping zones are considered in this model. Each set of small cells constitutes a so-called RoF cluster, as shown in Figure 4. The small cells can be integrated with the remote radio heads (RRHs), which are also connected to the central BS via high speed optical fiber or microwave links [12]. The APs of each cluster are controlled by a virtual base station (VBS) through optical fiber. The VBS is considered as a router of the RoF system. The system of load balancing and user transfer can be either distributed in each VBS or centralized in each virtual BS controller (VBSC)/virtual mobile switching center (VMSC). Each small cell is modeled by a multi-processor queue. Owing to the high density of small cells and in order to avoid the interference, some small cells are allowed to be inactive (idle mode) in the case of an interference occurring [13]. In this paper, the rates of the users (UEs) are limited by the core network. The proposed system model is assumed to accurately measure the user location from the user reference signals, and thus the location of each user is known [14].

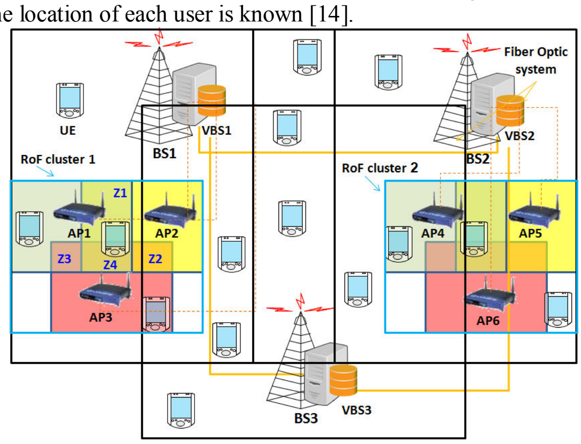


**Figure 1. Distributed system model.**

## 3. PROPOSED ALGORITHMS

The study consists of two main parts. In the first part, we introduce the LBA between the APs, which is first described in this section. In the second part, we present the user transfer algorithms. The following LBA between the APs with one of the upcoming two transfer algorithms are initialized. In other words, the LBA will be called by one of the two user transfer algorithms before or after transferring the extra users to the macrocells.

### 3.1 Load balancing algorithm (LBA)

In this section, we explain the LBA between the APs without any user transferred to the macrocells. The LBA, which is depicted in Figure 2, **first** starts checking the user density (ρ) within each cluster and comparing the density of the cluster with the highest density to the density threshold $\rho_{Th}$. If the user density does not exceed $\rho_{Th}$, the algorithm is stopped and it waits for the next trigger. Otherwise, the algorithm sets the throughput of each user, its zone and the tolerance parameter α, which will be determined later. After that, the algorithm calculates the throughput of each AP ($T_{AP(i)}$) as a summation of throughputs of all users (j) connected to the serving AP(i), as given by

$$T_{AP(i)} = \sum_{j=1}^{m(i)} T_{user(i,j)} \tag{1}$$

where m(i) is the number of users connected to AP(i). Next, the algorithm calculates the average network load (ANL) of the whole cluster as follows:

$$ANL = (T_{AP1} + T_{AP2} + \ldots T_{AP(n)}) / n \tag{2}$$

where n is the maximum number of APs. Meanwhile, the LBA determines the state of each AP by using the transfer policy. This policy verifies which AP must exclude a user (overloaded AP) and which one must include this user (underloaded AP). For that, two thresholds ($\delta_1$ and $\delta_2$) are needed. The upper threshold and the lower threshold are given by

$$\delta_1 = ANL + \alpha \times ANL, \;\; \delta_2 = ANL - \alpha \times ANL \tag{3}$$

According to the transfer policy, an underloaded AP can accept new users and handed-over users from an overloaded AP. While a balanced AP can only accept new users, an overloaded AP does not receive any new or handed-over users. Subsequently, the load balancing process will exclusively hand over the users from overloaded APs to underloaded APs.

With regard to the tolerance parameter α, the critical value of α is calculated before applying the LBA by setting the throughput of the most overloaded AP equal to $\delta_1$ as follows:

$$\alpha_{critical} = (T_{APmost-overloaded} - ANL) / ANL \tag{4}$$

Then, the result is divided by 10 to obtain the required value of α. In practice, the desired value of α can be empirically calculated so that the average value of α can be tuned based on the state and the location of the network.

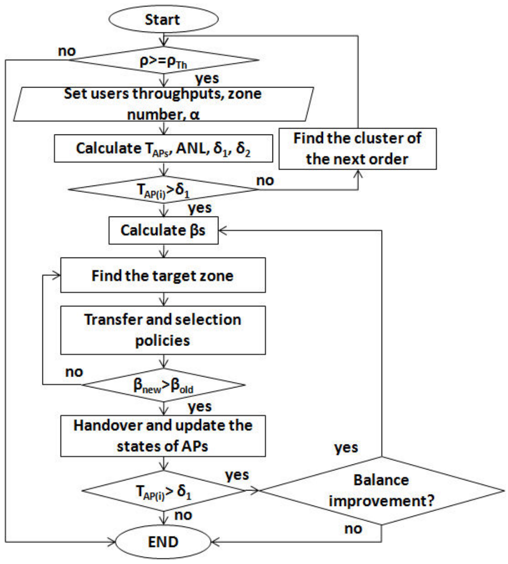


**Figure 2. Flowchart of the load balancing algorithm (LBA).**

In the **second step**, the algorithm checks if there is at least one overloaded AP within the cluster with the highest user density (cluster of first order). If not, the algorithm transits into the cluster of second or third order successively and rechecks the density condition. If this condition is not satisfied in these three clusters, the algorithm is stopped. Otherwise, the algorithm calculates Jain's fairness index (β) for each overlapping zone [15], which is determined as follows:

$$\beta = \frac{\left( \sum_{i=1}^{n} T_{AP(i)} \right)^2}{n \times \sum_{i=1}^{n} T_{AP(i)}^2} \tag{5}$$

where n is the number of small cells that overlap on the zone in question, i.e., each overlapping zone has its own β. When all the APs have exactly the same throughput, β is equal to one. Otherwise, β approaches 1/n, so $\beta \epsilon [1/n, 1]$. The **third step** is to apply the selection policy in order to determine the best candidate (BC) user to be handed over as follows. First, the difference between the selected overloaded AP and the ANL is calculated by

$$\Delta = T_{overloaded\ AP} - ANL \tag{6}$$

Of all the users located in the overlapping zone in question and connected to the chosen overloaded AP, the BC is the one for which the difference of the user throughput and delta has the smallest absolute value as follows:

$$BC_j = \left| T_{userj} - \Delta \right| \tag{7}$$

The **fourth step** is to determine the new β if the best candidate is handed over. This step is called the distribution policy. The aim of determining new β is to ensure that the expected handover will definitely improve the balance before doing the handover to avoid the ping-pong problem. Thus, the handover will be carried out if and only if $\beta_{new} > \beta_{old}$. Then, the algorithm selects this candidate and the handover decision occurs. Otherwise, the algorithm transits into the next target zone. It is one of the overlapping zones, which changes or not according to the selected load balancing scheme. After that, the algorithm repeats the last policies with the new target zone. The **fifth step** is to check again if there is still an overloaded AP within the selected cluster, and also if the balance improvement is still valid. If so, it evaluates the enhancement of the load balancing within the new target zone by updating all the values of β (βs) and so on. Otherwise, the algorithm is stopped and waits for the next trigger.

## 3.2 Transfer after algorithm (TAA)

The TAA is one of the algorithms that take care of the users, which should be transferred from the small cells to the macrocells, as depicted in Figure 3. This algorithm is composed of the following two stages. The first one is the balance stage achieved by the previous LBA. The second is the transfer stage, which is carried out after the balance stage. Therefore, the TAA has the same first steps of the LBA; however, when there are no more balance improvements within the APs, the transfer stage with new selection and transfer policies are initialized. In the **first step** of the transfer stage, the algorithm checks if at least one of the APs is overloaded, i.e., its throughput exceeds the transfer threshold $T_{capacity}$, which is the maximum allowed capacity for each AP. If not, the algorithm is stopped. Otherwise, the **second step** is to perform the new selection policy in order to determine the BC to be transferred by a vertical handover procedure as follows. First, the algorithm calculates the new delta as a difference between the most overloaded AP and $T_{capacity}$ as given by

$$\Delta_{new} = T_{most\text{-}overloaded AP} - T_{capacity} \tag{8}$$

Second, the best candidate value $BC_{(j)}$ is calculated for each user connected to the selected AP as a difference between the user throughput and the new delta as follows:

$$BC_{(j)} = T_{user(j)} - \Delta_{new} \tag{9}$$

Of all the users connected to the AP in question, the BC is the one for which the $BC_{(j)}$ has the smallest positive value. Otherwise, the BC is the one that has the smallest negative value, if all the values of $BC_{(j)}$ are negative. The transfer of the users from the chosen AP is repeated until the AP throughput becomes less than or equal to $T_{capacity}$ and as long as there are available users connected to this AP. In the **third step**, the algorithm determines the next most overloaded AP and then, it repeats the second step. When all the APs have been checked and there is not still any more transfer possibility, the TAA is stopped and waits for the next trigger.

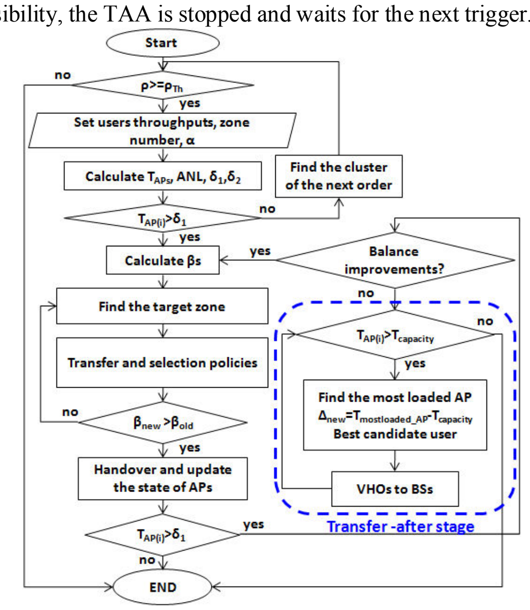


**Figure 3. Flowchart of the transfer after algorithm (TAA).**

## 3.3 Transfer before algorithm (TBA)

The TBA is similar to the TAA; however, the transfer stage is initialized as a first step for each AP that has a throughput exceeding $T_{capacity}$. When the throughputs of all the APs do not exceed $T_{capacity}$ any more or if there are no more available users to be transferred, the balance stage starts by calling the LBA, which continues the load balancing task as usual.

## 3.4 Active algorithm

The active algorithm is able itself to balance the load between the APs and to transfer the extra users to the macrocells without any help from the LBA unlike TAA or TBA. Actually, the active algorithm has a specific policy and is triggered for each new user entering into the network. This algorithm takes care of individual users by taking into account the user zone, the throughput of the user and the APs. This algorithm is composed of the following steps. The **first step** is to set the throughput of the new user and its zone. If the throughput of each AP is zero, i.e., this user is the first user that enters into the cluster. In this case, the user is transferred to one of the APs covering it based on the SNIR metric. Otherwise, the algorithm **in the second step** selects the best AP for this user. The selected AP is the least loaded AP so that if the throughput of this user is added to the throughput of this AP, the new AP throughput should not exceed $T_{capacity}$. If there is no AP satisfies this condition, the user is transferred to one of the macrocells. Once the AP is selected, the connection with the user is established. In the **third step**, the algorithm checks the density condition. The user density is calculated by considering the number of the users accepted in the UDN network and the number of the transferred users. If the user density is higher than or equal to $\rho_{Th}$, the algorithm is stopped. Otherwise, the algorithm sets the throughput of the next new user and so on. In practice, the density condition is not necessary to be verified, as this algorithm is always on standby and triggers for each new user. This condition is only imposed in order to compare the results of this algorithm to those of the previous algorithms with the same user density.

# 4. SIMULATION MODEL

To simplify, we consider two macrocells. Multiple small cells are covered by each macrocell. Each set of three overlapping small cells forms an RoF cluster, as shown in Figure 4. The load balancing is implemented at the cluster level. The tolerance parameter α is chosen to be 5%. The area of overlapping zones between each two small cells occupies about 25% of the total small cell area. The dimensions of each square small cell are 20x20 $m^2$ and the dimensions of each square macrocell are 0.5x0.5 $km^2$. The inter-sites distance is 15m. The user density (ρ) is on average equal to six users per small cell. Therefore, the density threshold is $\rho_{Th}$=18 users per cluster. Each user selects a specific throughput in the range from 0 to 350 Mbps [16]. The average throughput of each user is around 175 Mbps. Thus, the throughput for six users is around 1Gbps, which is the maximum allowed capacity for each AP, $T_{capacity}$. The total number of small square cells covered by each macrocell is 625 small cells. Hence, the maximum allowed capacity of each BS is 625 Gbps. Subsequently, the number of clusters, which can be covered by each BS is about 208 clusters, and the density of small cells per square kilometer is 2500 small cell/$km^2$. This density is greater than the value of $10^3$ small cell/$km^2$ used in the previous research studies, thus, the studied network is sufficiently overloaded. Due to the high density of UDN small cells, each new user is assumed to be covered at least by two or three APs.

# 5. PERFORMANCE CRITERIA

The considered criteria are Jain's index β and the standard deviation $\sigma_T$ of the throughputs of APs, which is given by

$$\sigma_T = \sqrt{\frac{(T_{AP1} - ANL)^2 + (T_{AP2} - ANL)^2 + ... + (T_{APn} - ANL)^2}{n-1}} \quad (9)$$

The index β4 is not always available for any configuration of small cells, as it depends on the network topology and on existing of a zone common between all the overlapping cells. In contrast, $\sigma_T$ is topology-independent and exists in any configuration of small cells. Moreover, $\sigma_T$ has a wider range than β, which is limited within [1/n, 1]. Thus, $\sigma_T$ gives a wider perspective for evaluating the results. Actually, β and $\sigma_T$ express the same state of balancing: the increment of the β value towards "1" leads to the decrement of the $\sigma_T$ value towards "0". Another criterion is the standard deviation of all the values of $\sigma_T$, STDEV($\sigma_T$). The latter represents an indicator about the change of traffic distribution among the APs during the different steps of the algorithm. It is not recommended to change the throughput of each AP suddenly, otherwise, the QoS of the users with real time applications might be affected. Other criteria are the handover rate (HOR), the rate of the transferred users (TR) and the balance improvement ratio (BIR), which is defined as the difference between the final value and the initial value of $\sigma_T$ divided by the initial value as given by

$$BIR = \left| \frac{\sigma_{T\,final} - \sigma_{T\,initial}}{\sigma_{T\,initial}} \right| \quad (10)$$

The best algorithm is the one that minimizes the required signaling and maximizes the load balancing. For that, we consider the balance efficiency (BE), the transfer efficiency (TE) and the overall efficiency (OE), which are respectively given by

$$Balance\ Efficiency\ (BE) = \sigma_T \times HOR \quad (11)$$

$$Transfer\ Efficiency\ (TE) = \sigma_T \times TR \quad (12)$$

$$Overall\ Efficiency\ (OE) = \sigma_T \times (HOR + TR) \quad (13)$$

# 6. PROPOSED APPROACHES

In order to accomplish the load balancing within the small cells, the following approaches are proposed:

## 6.1 Common Zone (CZ) Approach

In this approach, the load is only balanced by the users located in the common zone (CZ) between the three overlapping cells in the intersecting cell model. This zone is always the target zone. This approach is quick and simple, as it does not require much processing. In contrast, it is not very convenient in the case of UDN networks, since the user density is relatively low and the possibility to find many users located in the CZ might be limited.

## 6.2 Worst Zone (WZ) Approach

The load balancing in this approach is executed in the worst zone (WZ), which is the target zone with the smallest value of β(i). Thus, the CZ approach is less complicated than the WZ approach because the latter one must calculate the different β(i) to determine the WZ for each handover.

## 6.3 Mixed Approach (MA)

A hybrid approach that combines the CZ approach and the WZ approach. It starts balancing the load in the CZ and then, it transits into the WZ with or without returning to the CZ. Therefore, the target zone alternates between the CZ and the WZ. In this regard, we suggest the following five policies:

*6.3.1 2nd-AP policy* tries to hand over all the available users in the CZ as long as there are users of first order (connected to the most overloaded AP) and second order (connected to the next most overloaded AP). Then, it converts into the WZ approach.

*6.3.2 Early WZ policy* only executes one handover for a user of first order in the CZ and then transits early into the WZ approach.

*6.3.3 Persist 1st-users policy* only hands over the users of first order in the CZ before transiting into the WZ approach. Once there are more than one overloaded AP and the handovers for the first order users in the CZ are over by using the persist 1st-users policy, does the algorithm come back to the CZ or not? What are the potential policies in this case? To answer to these questions, two additional policies are proposed:

*6.3.4 Persist WZ policy* only hands over one user of second order in the CZ, after handing over all the users of first order by the persist 1st-users policy, then it converts into the WZ approach.

*6.3.5 Persist CZ policy* is opposite to the persist WZ policy, meaning that after handing over all the users of the first order, it only hands over one user by the WZ approach and then it transits into the CZ approach and tries to not return to the WZ approach.

On the other hand, in order to transfer the extra users to the macrocells, the following transfer approaches are proposed:

## 6.4 Passive approaches

We propose three approaches with intent to transfer the users to the macrocells. If the balance stage achieved by the LBA is carried out as a first step and then the transfer stage is performed as a next step, this approach is called the **transfer after (TA)**

**approach**. Presumably UDN networks offer calls with less expensive cost and provide better QoS, particularly the latency, hence, the preferred approach is the one that transfers the smallest number of users to the macrocells. Alternatively, if the transfer is achieved as a first step, this approach is called the **transfer before (TB) approach**. Both approaches are named passive, as they are only triggered when the density condition is satisfied.

## 6.5 Active approach (AA)

The AA approach, opposite to the passive approaches, is always on standby and ready to be triggered each time a new user. This approach depends on the throughput of the user and the APs, and also on the zone. When an AP is selected to include the new user and the throughput of this AP will not exceed $T_{capacity}$ if it accepts this new user, this user is accepted by this AP. Otherwise, the algorithm transfers this user. This process is repeated for each new user until the user density of the chosen cluster reaches $\rho_{Th}$.

# 7. RESULTS ANALYSIS

To clarify, we first discuss the results of all the previous approaches for an example of applications. Consider a cluster with three intersecting small cells and four overlapping zones (Z1, Z2, Z3 and Z4) covered by two macrocells, as shown in Figure 4. Four values for Jain's index (β1, β2, β3 and β4) are assumed with number of overlapping zones as n=2, 2, 2 and 3, respectively. In this cluster, 18 users exist and each user is represented by its number j. In this model, Z1 is the overlapping zone between the two cells of AP1 and AP2. The following users are located in Z1: UE3, UE4 and UE9. Similarly, Z2 is the overlapping zone between the two cells of AP2 and AP3. UE1, UE2, UE6, UE8 and UE14 are located in Z2. UE11 and UE18 are located in Z3, which is the overlapping zone between the two cells of AP1 and AP3. Finally, UE5, UE7, UE10, UE12, UE13, UE15, UE16 and UE17 are located in Z4, which is the CZ between all the cells.

Applying the TB approach using the WZ approach, all the APs became balanced or underloaded according to $T_{capacity}$ and the APs can thus accept new users. However, before applying this approach, only one AP could accept new users. Hence, the resource utilization is highly increased and the throughputs are redistributed better among the small cells.

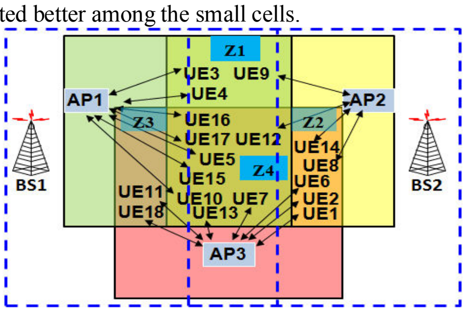


**Figure 4. A cluster of three intersecting small cells.**

Figure 5 (a), Figure 5 (b), Figure 6 (a), and Figure 6 (b) show the throughputs of all the APs, $\sigma_T$, β and STDEV($\sigma_T$) versus the transferred and handed-over users. The throughputs of the APs are finally located inside the desired balance range [$\delta_1$, $\delta_2$]. While β tends towards 1, $\sigma_T$ goes to zero. STDEV($\sigma_T$) increases according to the throughputs of the transferred users and the handovers. This increase confirms that the handovers and the transfer processes are achieved for the users with the highest throughputs. Hence, the throughputs of the APs greatly change in the TB approach. This can be considered the biggest drawbacks of this approach.

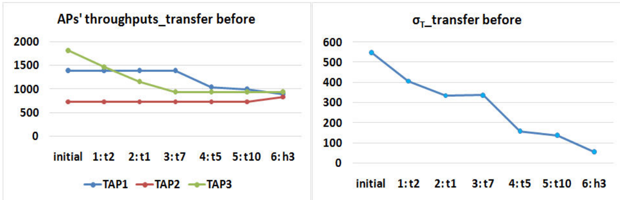


**Figure 5. The throughputs of APs (a) and $\sigma_T$ (b).**

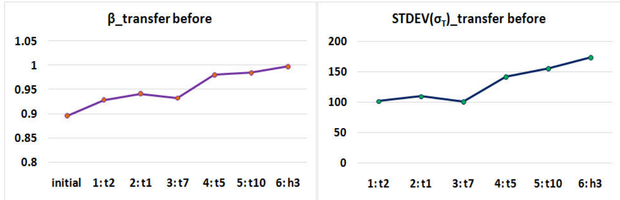


**Figure 6. β (a) & STDEV($\sigma_T$) (b).**

Applying the TA approach using the WZ approach, the final states of the APs become better than in the TB approach. In addition, the number of the transferred users becomes smaller than that of the TB approach at the expense of increasing the handovers. In fact, the TAA is based on first handing over the users and then, transferring the extra users to the BSs. In that way, it guarantees the balance among the small cells and then transits into the transfer task. Moreover, in the case of TA approach, we found that STDEV($\sigma_T$) decreases versus the handovers and the transferred users. Because the TA approach selects the users with the lowest throughputs, i.e., the traffic redistribution among the APs is carried out more smoothly than in the TB approach.

Applying the AA approach, the balancing results outperform those in the case of the TB approach at the expense of more signaling load caused by the frequent triggering of the active algorithm for each new user. Moreover, STDEV($\sigma_T$) is smaller than in the passive approaches. This value decreases smoothly due to the smaller change of the AP throughput caused by each new user. On the other hand, we found that the criteria of the AA approach are more dependent on the throughput of the new user and to which AP the new user will be connected. Furthermore, the curves of $\sigma_T$, β and STDEV($\sigma_T$) are more or less similar to the TA approach. However, the APs throughputs curves start from zero and end up inside the desired balance rang [$\delta_1$, $\delta_2$]. Additionally, the $\sigma_T$ curve fluctuates much more than in the passive approaches. In addition, the AA approach does not require any handover between the APs.

In the following, the general results are discussed. We found that the TA approach achieves the best **balancing results ($\sigma_T$)**, particularly using the MA approach (the average value of the MA policies). In fact, the MA approach finely balances the load among the small cells, as it has the maneuvering capability between the CZ and the other WZs. On the other hand, we observed in the case of TB approach that sometimes one AP is kept slightly overloaded. As many users are transferred to the macrocells early on, no more BCs are available to extensively balance the APs. Besides, the worst balancing results are noticed using the CZ approach, as it hands over a smaller number of users than the MA approach. Hence, a tradeoff between the balance improvements and the number of handed-over users exists. Additionally, all the MA policies illustrate identical results in the case of TB approach. This means that these policies are more efficient in the case of TA approach. The MA approach enhances the balance (on average) by 82.53% and 60.97% compared to the WZ and the CZ approaches, respectively. Moreover, the TA approach improves the balancing results more than the TB approach and the AA approach by 49.53% and 9.70%, respectively. Comparing to the case without transfer, while the TA and the AA approach improve the balancing results by 41.28% and 34.97%, respectively, the TB one decreases the balancing results by 16.33%. Note that similar results are noticed based on the β index.

On the other hand, the TA approach achieves the same **handover rate (HOR)** as the case without transfer, even though it illustrated excellent balancing results. This is one of the best advantages of the TA approach. Regarding the **rate of the transferred users (TR)**, the TA approach shows the smallest TR. In contrast, the TB approach achieves the highest TR. Since the TB approach transfers before balancing, this forces it to transfer many more users to reach the intended balance for each AP exceeding $T_{capacity}$ regardless of the state of balance between the small cells.

On the other hand, the TA approach achieves the best **balance improvement ratio (BIR)**. This ratio reaches 97.94% using the MA approach. The latter is better by 5.01% than the case without transfer. Alternatively, the AA approach leads to the worst BIR. In fact, as the specific policy of the AA is already efficient in balancing the load, it cannot be improved more by transferring the users. Regarding the **balance efficiency (BE)**, it is improved compared to the case without transfer. The TA approach using the MA approach is the best option to balance the load. In this case, the BE increases by 52.44% and 76.44%, respectively compared to the TB approach and to the case without transfer. Considering the **transfer efficiency (TE)**, the TB approach leads to the worst TE, the TA approach is again the best option with the MA approach. On the other hand, the TA approach using the MA approach reveals the best **overall efficiency (OE)**. Its OE using the MA approach is better by 80.34%, 50.41% and 51.32% than that of the TB approach, the AA approach and the case without transfer, respectively. As the purpose is to balance the load among the small cells and simultaneously to reduce as much as possible the number of the users transferred to the macrocells, this purpose is met by using the TA approach with the MA approach.

# 8. CONCLUSION

The importance of the load migration mechanism within the UDN networks is studied. This mechanism can be found on the load balancing approaches in order to redistribute the uneven traffic throughputs among the small cells and transfer the extra users to the macrocells. The load balancing algorithm would be a promising solution to enhance the performance of UDN networks. In this context, different approaches are proposed. While the mixed approach shows the best balancing results, the worst zone approach is the most efficient one in the case without user transfer. To transfer the users to the macrocells, several user transfer approaches are proposed. An active approach (AA) is not preferred, as it should be constantly on standby for balancing the network by user-by-user way and transfer the extra users. It may result in a network congested by signaling. Consequently, it is better to transfer the users based on the passive approaches. In this regard, the transfer after (TA) approach and the transfer before (TB) approach are suggested. As the intended purpose is to keep as many users as possible within the UDN network, the TA approach using the MA approach is the most efficient approach. Perspective works deal with the device-to-device and machine-type-communications using other load balancing methods, which take care of the end-to-end delay for the users. In this paper, the base stations are permanently considered available to accept the transferred users; however, the grade of service and the load balancing between the macrocells can be also highlighted.

# 9. REFERENCES


[1] Zhen He, H. Xu and Jianping Wang, "Non-cooperative differential game based load balancing algorithm in Radio-over-Fibre system," *International Conference on Cyberspace Technology (CCT 2014)*, Beijing, 2014, pp. 1-4.

[2] Cooper A J. 'Fiber/Radio' for the provision of cordless/mobile telephony services in the access network[j]. Electron. Lett., -2054:(24)1990,26.

[3] Haitao Xu, Zhen He and Xianwei Zhou, "Load Balancing Algorithm of Ultra-Dense Networks: a Stochastic Differential Game based Scheme," KSII Transactions on Internet and Information Systems, vol. 9, no. 7, pp. 2452-2467, 2015.

[4] Balachandran A, Bahl P, Voelker G M. Hotspot congestion relief and service guarantees in public-area wireless networks[J]. ACM SIGCOMM Computer Communication Review, 2002, 32(1): 59-59.

[5] V. Aleo Load Distribution In IEEE 802.11 Cells, Master of Science Thesis, KTH, Royal Institute of Technology, 2003.

[6] Y.T. Wang. A Fuzzy-Based Dynamic Channel Borrowing Scheme for Wireless cellular Networks. in: Proceedings 57th IEEE Conference on Vehicular Technology, USA: IEEE Press,2003,3:1517-1521

[7] Patra S S M, Roy K, Banerjee S, et al. Improved genetic algorithm for channel allocation with channel borrowing in mobile computing[J]., IEEE Transactions on Mobile Computing, 2006, 5(7): 884-892.

[8] Hanly S V. An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity[J]. Selected Areas in Communications, IEEE Journal on, .1340-1332 :(7)13 ,1995

[9] Yunting Wang, Xiaodong Xu, Yaqi Jin: *QoS constraint optimal load balancing for Heterogeneous Ultra-dense Networks*. WPMC 2016: 317-323.

[10] I. Elgendi, K. S. Munasinghe and A. Jamalipour, "Traffic offloading for 5G: L-LTE or Wi-Fi," 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Atlanta, GA, 2017, pp. 748-753.

[11] M. Ding, D. Lopez Perez, G. Mao, ‘’ A New Capacity Scaling Law in Ultra-Dense Networks, ‘’ arXiv: 1704.00399V1 [cs.NI], 3 Apr 2017.

[12] M. Peng, Y. Li, Z. Zhao, and C. Wang, “System architecture and key technologies for 5G heterogeneous cloud radio access networks,” IEEE Network, vol. 29, no. 2, pp. 6–14, Mar./Apr. 2015.

[13] H. Claussen, I. Ashraf, and L. T. W. Ho, “Dynamic idle mode procedures for femtocells,” Bell Labs Tech. J., vol. 15, no. 2, pp. 95–116, .2010.

[14] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen and M. Valkama, "High-Efficiency Device Positioning and Location-Aware Communications in Dense

5G Networks," in *IEEE Communications Magazine*, vol. 55, no. 8, pp. 188-195, 2017.

[15] Miaona Huang, Suili Feng, and Jun Chen, A Practical Approach for Load Balancing in LTE Networks, Journal of Communications Vol. 9, No. 6, pp. 490-497 June 2014.

[16] Petteri Kela, Continuous Ultra-Dense Networks, A System Level Design for Urban Outdoor Deployments, book 1799-4942 (electronic), Aalto University publication series DOCTORAL DISSERTATIONS 86/2017.